STIC-ILL

AHL
QH426 .G46

From: Davis, Minh-Tam
Sent: Tuesday, January 28, 2003 3:28 PM
To: STIC-ILL
Subject: reprint request for 09/807201

Paoloni-Giacobino, A, 1997, Genomics, 44(3): 309-320.
Thank you.
MINH TAM DAVIS
ART UNIT 1642, ROOM 8A01, MB 8E12
305-2008

GENOMICS 44, 309–320 (1997)
ARTICLE NO. GE974845

Cloning of the TMPRSS2 Gene, Which Encodes a Novel Serine Protease with Transmembrane, LDLRA, and SRCR Domains and Maps to 21q22.3

Ariane Paoloni-Giacobino,* Haiming Chen,* Manuel C. Peitsch,† Colette Rossier,* and Stylianos E. Antonarakis*,‡,[1]

*Laboratory of Human Molecular Genetics, Department of Genetics and Microbiology, Geneva University Medical School, Geneva; †Glaxo Institute for Molecular Biology, Geneva; and ‡Division of Medical Genetics, Cantonal Hospital of Geneva, 1211 Geneva, Switzerland

Received March 24, 1997; accepted June 6, 1997

To contribute to the development of the transcription map of human chromosome 21 (HC21), we have us d exon trapping from pools of HC21-specific cosmids. Using selected trapped exons, we have identified a nov l g ne (named TMPRSS2) that encodes a multimeric protein with a serine protease domain. The TMPRSS2 3.8-kb mRNA is expressed strongly in small int stine and weakly in several other tissues. The full-length cDNA encodes a predicted protein of 492 amino acids that contains the following domains: (i) A serine proteas domain (aa 255–492) of the S1 family that probably cleaves at Arg or Lys residues. (ii) An SRCR (scaveng r receptor cysteine-rich) domain (aa 149–242) of group A (6 conserved Cys). This type of domain is involved in the binding to other cell surface or extracellular molecules. (iii) An LDLRA (LDL receptor class A) domain (aa 113–148). This type of domain forms a binding site for calcium. (iv) A predicted transmembran domain (aa 84–106). No typical signal peptide was recognized. The gene was mapped to 21q22.3 between markers ERG and D21S56 in the same P1 as MX1. The physiological role of TMPRSS2 and its involvement in trisomy 21 phenotypes or monogenic disorders that map to HC21 are unknown. © 1997 Academic Press

INTRODUCTION

Human chromosome 21 (HC21) is the smallest chromosome, with a long arm (21q) of around 40 Mb, containing approximately 600–1000 genes (reviewed in Antonarakis, 1993), and a short arm (21p) of around 10–15 Mb, which is highly homologous to those of the other four human acrocentric chromosomes. To date, about 75 HC21 genes have been cloned and partially characterised [Genome DataBase, http://gdbwww.gdb.org, and SWISS-PROT, http://www.expasy.ch]. Trisomy for human chromosome 21 is the most common chromosomal abnormality at birth, leading to the phenotypes known as Down syndrome (Epstein, 1989). In addition, the loci for several monogenic disorders have been mapped to HC21. Dense linkage maps and almost complete physical maps of 21q have already been obtained and are now extensively used for the characterization of HC21 genes and the efforts to determine the nucleotide sequence of HC21. The cloning and characterization of HC21 genes are a necessary step for the understanding of Down syndrome and the molecular etiology of monogenic disorders mapping on this chromosome.

In our laboratory, systematic exon-trapping experiments have been performed to identify portions of HC21 genes, clone and characterize the corresponding full-length cDNAs and genes, and participate in the international effort to create a transcription map of HC21 (Cheng *et al.*, 1994; Peterson *et al.*, 1994; Tassone *et al.*, 1994; Lucente *et al.*, 1995; Chen *et al.*, 1996). We report here the cloning of a novel serine protease gene (TMPRSS2), which is expressed mainly in the small intestine, but also in lower levels in several other tissues, and which maps to 21q22.3. The predicted polypeptide of TMPRSS2 also contains a transmembrane domain, a scavenger receptor cysteine-rich (SRCR) domain, and an LDL receptor class A (LDLRA) domain, and it probably belongs to the type II integral membrane proteins. The TMPRSS2 gene is homologous to, but different from, the human enteropeptidase gene, which maps to a different region of HC21 (21q21).

MATERIALS AND METHODS

Exon Trapping

Pools of chromosome 21-specific cosmids from the LL21NCO2 library (kindly supplied by P. de Jong) were used in exon-trapping

Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. U75329 (cDNA) and X88229, X88228, X88321, X88043, and X88047 (trapped exons).

[1] To whom correspondence should be addressed at Division de Génétique Médicale, Centre Médical Universitaire, 1 rue Michel-Servet, 1211 Genève 4, Switzerland. Telephone: 41-22-7025707. Fax: 41-22-7025706. E-mail: Stylianos.Antonarakis@medecine.unige.ch.

0888-7543/97 $25.00

experiments (Buckler *et al.*, 1991; Church *et al.*, 1994; Gibco BRL Manual 18449-017). *Eco*RI- and *Pst*I-digested cosmids were subcloned into pSPL3 vector, and plasmid DNA was used to transfect Cos7 mammalian cells using lipofectACE (Gibco BRL). Total RNA was isolated from Cos7 cells 24 h after transfection, cDNA was synthesized, and PCR products were subcloned into pAMP10 vector by UDG (uracil DNA glycosylase) cloning. After elimination of cryptically spliced, pSPL3-derived clones by oligonucleotide screening, the inserts of individual pAMP10 clones were subjected to nucleotide sequencing on an ABI373A automated sequencer by dideoxy terminator fluorescence method using *Taq* polymerase. Nucleic acid and amino acid homologies of the resulting sequences were analyzed through BLASTN and BLASTX searches of the nonredundant database (Altschul *et al.*, 1990).

Cloning of TMPRSS2 cDNA

The 216-bp PCR product derived from trapped exon HMC26A01 with oligonucleotide primers (26A01A, 5′-GCCTGCGGGGTCAACTTGAAC-3′, and 26A01B, 5′-GGCGGCTGTCACGATCCACTC-3′) was used as a probe to screen approximately 500,000 clones of a human heart λgt10 cDNA library (Clontech HL3026a). One positive clone (APG1) was isolated, and the 2.4-kb insert was subcloned into the pAMP10 vector and sequenced in both directions using standard oligonucleotide walking protocols for the ABI373 automated sequencer. The nucleotide sequence was verified using RT-PCR products from intestine poly(A)$^+$ mRNA.

Chromosomal Mapping

Two independent methods were used to assign TMPRSS2 to a human chromosome. First, PCR amplification of the trapped exon HMC26A01 with specific oligonucleotide primers (26MAP1, 5′-GAGGCTTCTGCAGCTTCATC-3′, and 26MAP2, 5′-CAATCCATGGCATTGGACGG-3′) was performed on the genomic DNA from a panel of somatic cell hybrids with defined segments of HC21. Second, the insert of the initial trapped exon HMC26A01 was used to probe high-density filters of cosmids from the HC21-specific LL21NC02 library. Finally, PCR amplification using either oligonucleotide primers 26 MAP1 and 26 MAP2 or 26A01A and 26A01B was used on DNAs from a panel of HC21-derived YACs.

5′- and 3′-RACE (Rapid Amplification of cDNA Ends)

To obtain the 5′ end of the TMPRSS2 cDNA, 5′-RACE was performed on human small intestine cDNA. From 1 μg of poly(A)$^+$ RNA (Clontech 6547-1) cDNA was made with the Marathon cDNA Amplification kit (K-1802-1), and 5′-RACE using nested PCR primers was carried out with the enzyme *Taq* Expand High Fidelity (Boehringer Mannheim) according to the manufacturer's protocol. The gene-specific primers were 26A01B (see above) and AP26BB (5′-CCGCTGTCATCCACTATTCC-3′). In two different experiments the same PCR product of 670 bp was generated and subjected to nucleotide sequencing. 3′-RACE was carried out using gene specific primers AP26G (5′-GGTTCTGGCTGTGCCAAAGC-3′) and AP26K (5′-GTCTGGCTTTGGCACTCTCTGC-3′), and a PCR product of approximately 2.0 kb was generated.

Northern Blot Analysis

The cDNA clone APG1 containing the complete coding sequence was used to probe two Northern blots, each containing poly(A)$^+$ RNA from eight human adult tissues (Clontech 7759-1, Clontech 7760-1), and one containing four fetal tissues (Clontech 7756-1). Northern Blot analysis was performed using standard protocols, with high-stringency washing. A control hybridization using a human actin probe was used for determination of the amount of RNA loaded in these Northern blots.

Comparative Protein Modeling

The sequences of both LDLRA and protease domains of TMPRSS2 were submitted to the SWISS-MODEL automated comparative protein modeling server (Peitsch, 1995, 1996). The models were made as follows:

LDLRA domain. SWISS-MODEL could not automatically provide a 3D structure of this domain since the degree of identity with the most similar sequence of known 3D structure was less than 30%. Using BLAST (Altschul *et al.*, 1990), we identified the Brookhaven Protein Data Bank entry 1LDL (NMR structure of the LDLR1 domain) (Daly *et al.*, 1995) as the suitable modeling template. We then aligned the TMPRSS2 LDLRA domain with the sequence of 1LDL and submitted the sequence alignment to SWISS-MODEL using the Optimise mode.

Serine protease domain. This domain was modeled using the First Approach mode of SWISS-MODEL, which provides fully automated template identification and multiple sequence alignment prior to model building. Chymotrypsin (P17538) was identified as a suitable modeling template. The template and TMPRSS2 protease sequences were automatically aligned and the model generation proceeded to the end without human intervention. Sequence to structure fitness analysis using both 3D–1D profiles (Lüthy *et al.*, 1992) and Prosall (Sippl, 1993) did not show any obvious discrepancies. The coordinates of both the LDLRA and the serine protease domain of TMPRSS2 can be found in the SWISS-MODEL Repository (http://www.expasy.ch/swissmod/swmr-top.html).

RESULTS

Exon Trapping Identified a Clone with Homology to Human Proteases

To clone partial gene sequences from human chromosome 21 we have used pools of cosmids (from the LL21NCO2-Q library) in an exon-trapping experiment and have identified more than 550 different potential exons (Chen *et al.*, 1996). One trapped sequence HMC26A01 (GenBank X88229) of 216 bp showed a strong homology to a large list of serine proteases from human and other species. BLASTX analysis, for example, revealed a 55% amino acid identity to human prostasin (GenBank L41351; $P = 1.3e{-}15$). Other representative homologies included human elastase (P08218), *Erinaceus europaeus* plasminogen (U33171), and pig human coagulation factor IX (P16293). Because this HMC26A01 trapped sequence was probably derived from a undescribed human serine protease, we set out to clone and initially characterize the full-length cDNA of the corresponding human gene.

Isolation of Full-Length TMPRSS2 Coding Sequences

Clone HMC26A01 was used to screen approximately 500,000 clones of a human heart λgt10 cDNA library (this library was chosen because of the expression pattern in Northern blots; see below). One positive clone (APG1), containing a 2.4-kb-long insert, was obtained, subcloned into the pAMP10 vector, and subjected to nucleotide sequence. 5′-RACE from intestinal mRNA (again chosen because of the expression pattern) using oligonucleotides close to the 5′ end of the APG1 clone extended the 5′UTR sequence by about 150 nucleotides. Sequence analysis from both strands revealed an open reading frame of 492 amino acids starting from the most N-terminal methionine codon. The 3′UTR from the original clone APG1 was approximately 0.95 kb. Figure 1 shows the complete nucleotide

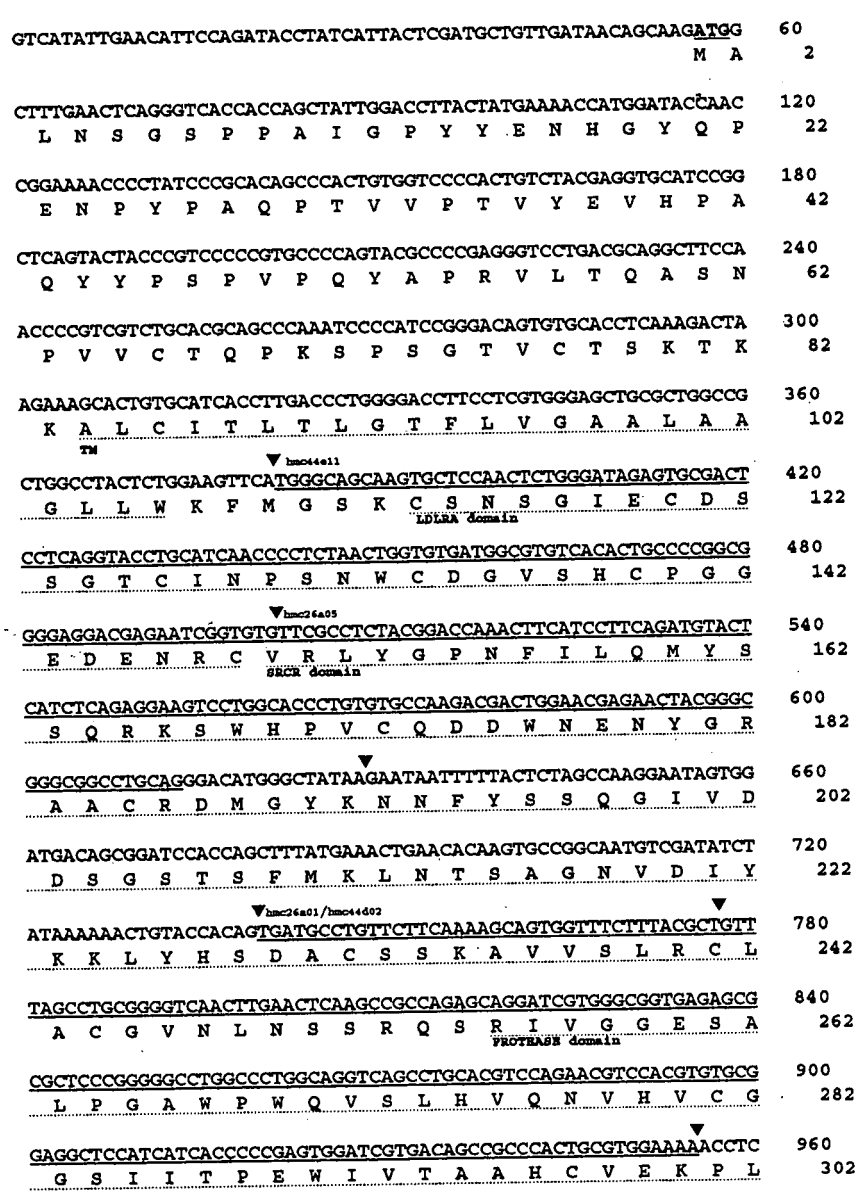


FIG. 1. Nucleotide and predicted amino acid sequence of TMPRSS2 (GenBank Accession No. U75329). The potential initiation methionine codon and the translation stop codons are bold and underlined. The trapped sequences are underlined (for example the trapped sequence HMC26A01 extending from nucleotide 740 to 955). The different domains of the predicted polypeptide are dotted underlined (for example the SRCR domain extends from amino acid residue 148 to 242). The locations of the introns are shown with arrows. For the sequences of intron/exon junctions see Fig. 2.

```
...agggcacctctcttctgttttctctgcaag /TGGGCAGCAA.....AATCGGTGTG/ gtgagtcagccttaaccttgggaagggact...
                                  G110                        V149

...aactcatggataatcctccctctcgtgcag /TTCGCCTCTA.....TGGGCTATAA/ gtgagtatggggcagcacccgccgagtgac...
                                  R150                        K191

...cgtgaccagaatttcccgttctttttgcag /TGATGCCTGT.....TCTTTACGCT/ gtataggtaagttcatctggagtccccctt...
                                  D229                        C241

...ctgagatactgagtccttcttctctcccag /ACCTCTTAAC.....ACTTTCAACG/ gtacgtgtggctcaggcttggcaagcaggt...
                                  P301                        D359

...ggctcactgtgttttctcttcttgaaacag /ACCTAGTGAA.....GAGGAGAAAG/ gtgaggctgctcctgggcacacaggactgc...
                                  L360                        G391

...tgggagctcaacaagtctccctgtccttag /GGAAGACCTC.....TTCTTGCCAG/ gtaattcaacatttttattctacctttggtc...
                                  K392                        Q438

...ctgctctctgtaccttgctgtgtcccacag /GGTGACAGTG.....ATGAAGGCAA/ gtaactatcctgtcctccttctgactgtgtt...
                                  G439                        N491

...cactttttttttcctatttgaacaggcag /ACGGCtaatccacatggtcttcgtccttgacgtcgp(3UTR)...
                                  G492 *
```

FIG. 2. Intron/exon junctions of the TMPRSS2 gene as determined by comparison of the cDNA sequence to the publicly available sequences of the human P1 clone 35-H5-C8 (Martin *et al.*, 1994; Genbank Accession Nos. L35675–L35682).

and predicted amino acid sequence of TMPRSS2. This cDNA was verified by RT-PCR amplifications from intestinal RNA using pairs of oligonucleotide primers from the cDNA sequence. Interestingly, no ESTs identical to portions of the TMPRSS2 cDNA sequence were identified in the dbEST database of GenBank (search of February 18, 1997). A number of additional exons from the Chen *et al.* (1996) study were identical to portions of the TMPRSS2 cDNA, including HMC44E11 (GenBank X88043), HMC26A05 (GenBank X88228), HMC19A07 (GenBank X88321), and HMC44D02 (GenBank X88047).

Intron/Exon Junctions

Homology searches with sequences available in the public databases revealed identity of discontinuous regions of the TMPRSS2 cDNA with portions of human P1 clone 35-H5-C8 which was sequenced by Martin and co-workers (Martin *et al.*, 1994; GenBank Accession Nos. L35675–L35682). The comparison of the cDNA sequence of TMPRSS2 with the genomic sequence of human P1 revealed intron/exon junctions that are shown in Fig. 2. Not all such junctions are reported in the figure since the sequence of the entire P1 clone was not available in the public databases. It is likely that there are additional introns 5′ to codon 110 and between codons 191 and 229 and codons 241 and 301.

Mapping of TMPRSS2 to Chromosome 21

PCR amplification was performed with oligonucleotide primers 26MAP1 and 26MAP2 on genomic DNA from rodent–human somatic cell hybrids that contained either single human chromosomes (NIGMS 2; Drwinga *et al.*, 1993) or specific segments of HC21 (Patterson *et al.*, 1993). The expected 155-bp PCR product was present in somatic cell hybrids WAV17, E7b, 725, 2Fur1, R50-3, GA9-3, 9528C-1, 1881C-13b, 8q−, ACEM 2-10d, JC6A, and 1×4; in contrast, somatic cell hybrids 21q+, 6918-8al, and MRC2-G did not show amplification (data not shown). These data localized this human protease to the region 21q22.3 between markers ERG and D21S56 (Fig. 3).

We used exon HMC26A01 to probe a subset of the cosmid library LL21NC02. One cosmid, Q20A3, was identified as positive. PCR on this cosmid with the same primers 26MAP1 and 26MAP2 produced the expected 155-bp fragment, confirming that Q20A3 contained this exon of TMPRSS2 gene. Yeast DNA from 79 YAC clones, chosen to cover almost all of HC21 (Chu-

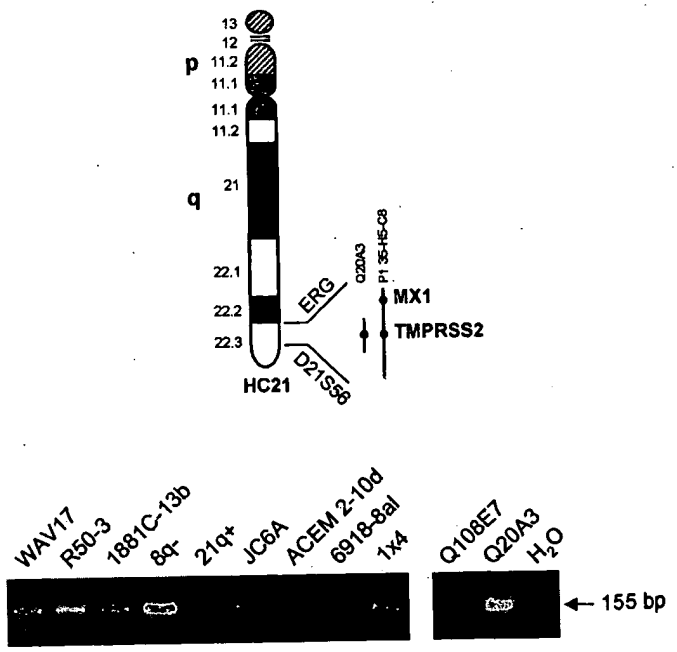


FIG. 3. Schematic representation of the mapping position of the TMPRSS2 gene on chromosome 21 as resulted from PCR amplification of somatic cell hybrids and sequence identities with a chromosome 21 P1 clone (see Results). Representative results from PCR amplification using oligonucleotide primers 26MAP1/26MAP2 (see text) are also shown.

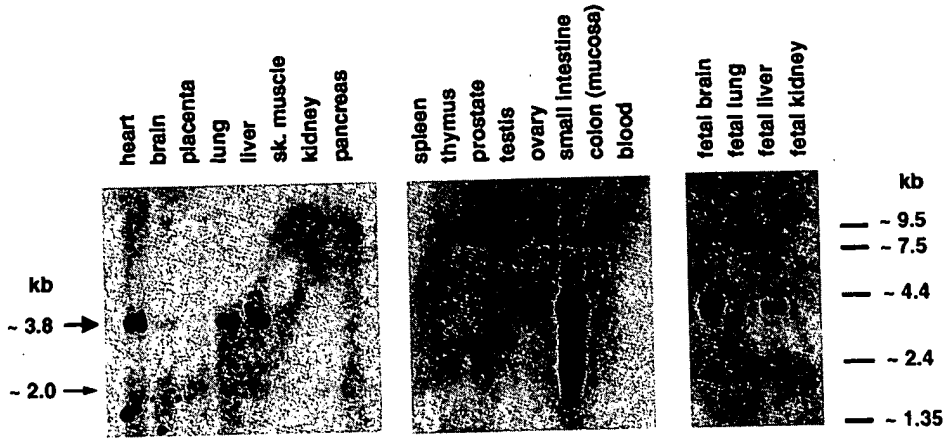


FIG. 4. Northern blot analysis using the TMPRSS2 cDNA as hybridization probe. The RNA filters are from Clontech (Cat. Nos. 7750-1, 7760-1, 7759-1, and 7756-1) and contain 2 μg of poly(A)$^+$ mRNA per tissue indicated. The thick arrow shows the 3.8-kb mRNA species, while the thin arrow depicts the faint 2.0-kb mRNA.

makov *et al.*, 1992), was used for PCR amplification with the two pairs of oligonucleotide primers 26MAP1–26MAP2 and AP26G (5′-GGTTCTGGCTGTGCCAA-AGC-3′)–AP26H (5′-CCAATGTGCAGGTGGAGACC-3′) in the 3′UTR region. No positive YACs were identified. Many single YACs in 21q22.3 from the collection of Chumakov *et al.* (1992) were also tested by PCR with these primers and no amplification was observed. The absence of positive YACs for this human TMPRSS2 gene suggests either that the HC21 contig (Chumakov *et al.*, 1992) in the region between markers ERG and D21S56 contains at least one gap or that the YAC clones available to our laboratory have accumulated deletions.

As described above, discontinuous regions of the TMPRSS2 cDNA were identical to portions of human P1 clone 35-H5-C8, which was sequenced by Martin and co-workers (Martin *et al.*, 1994; GenBank Accession Nos. L35675–L35682). This P1 also contained gene MX1, which maps to 21q22.3 in the interval between ERG and D21S56 (Fig. 3). Therefore, this sequence identity of TMPRSS2 with portions of P1 35-H5-C8 is in agreement with the mapping position obtained using the somatic cell hybrids.

Northern Blot Analysis

The insert of cDNA clone APG1 was used as a probe against three filters containing 2 μg of poly(A)$^+$ RNA from 16 human adult tissues and 4 human fetal tissues. A hybridization signal corresponding to an mRNA species of approximately 3.8 kb was detected (Fig. 4). The difference between the 2.4-kb cDNA clone APG1 and the 3.8-kb RNA species detected in the Northern blot is probably due to the continuation of the 3′UTR downstream of the end of clone APG1. 3′-RACE from intestine RNA using oligonucleotides from clone APG1 (oligonucleotide primers AP26G, see above, and AP26K 5′-GTCTGGCTTTGGCACTCTCTGC-3′) revealed a PCR product of approximately 2.0 kb, which corresponds to a mRNA length of 3.8 kb, compatible with the results of the Northern blot analyses (data not shown). The highest level of expression was observed in small intestine, but this gene is also expressed in human adult heart, placenta, lung, thymus, and prostate and in fetal brain and liver. Another weakly hybridizing mRNA species of 2.0 kb was also observed in several tissues. This could be due to alternative splicing, utilization of different transcription start sites and polyadenylation signals, overlapping transcripts, or, most likely, cross-hybridizing transcripts with sequence homologies with TMPRSS2. A human actin probe was used to control the amount of RNA loaded (data not shown). The expression of the TMPRSS2 gene appears to be developentally regulated since there is strong expression in fetal brain but very little expression in adult brain. In addition, in the lung, expression is high in the adult tissue but low in the fetal tissue.

Type II Transmembrane Protein

Protein prediction programs, which predict transmembrane domains, including http://ulrec3.unil.ch/software/TMPRED_form.html (Hofmann and Stoffel,

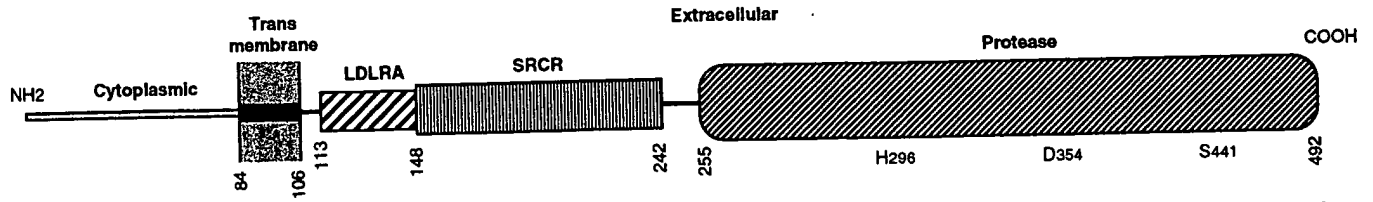


FIG. 5. Schematic representation of the different domains of TMPRSS2. Numbers correspond to codons of the full-length cDNA shown in Fig. 1. For description of the domains see text.

a

```
 1 TMPRSS2       5  C-SNSGIECDSS-GTCINPSNWCDGVSHCPGGEDEN---RC
 2 EK1         199  C-PPDSRLCADA-LKCIAIDLFCDGELNCPDGSDEDNK-TC
 3 EK4         659  C-KEDNFQCKD--GECIPLVNLCDGFPHCKDGSDE---AHC
 4 LDLR1         6  C-ERNEFQCQD--GKCISYKWVCDGSAECQDGSDES-QETC
 5 LDLR2        47  C-KSGDFSCGGRVNRCIPQFWRCDGQVDCDNGSDEQ---GC
 6 LDLR3        88  C-SQDEFRCHD--GKCISRQFVCDSDRDCLDGSDE---ASC
 7 LDLR4       127  C-GPASFQCNS--STCIPQLWACDNDPDCEDGSDEW-PQRC
 8 LDLR5       176  C-SAFEFHCLS--GECIHSSWRCDGGPDCKDKSDE---ENC
 9 LDLR6       215  C-RPDEFQCSD--GNCIHGSRQCDREYDCKDMSDE---VGC
10 LDLR7       255  CEGPNKFKCHS--GECITLDKVCNMARDCRDWSDEP-IKEC
11 L34049a    1066  C-SPSAFACVRG-GQCIPGQWHCDRQNDCLDGSDEQ---NC
12 L34049b    1188  C-TSAQFKCADG-SSCINSRYRCDGVYDCRDNSDEA---GC
13 U13637a    1074  C-SPSQFACHSG-EQCVDKERRCDNRKDCHDHSDEQ---HC
14 U13637b    1283  C-SIYEFKCRSG-RECIRREFRCDGQKDCGDGSDEL---SC
15 P07358      122  C---EGFVCAQT-GRCVNRRLLCNGDNDCGDQSDEA---NC
16 U60975     1471  C-DRFEFECHQP-KTCIPNWKRCDGHQDCQDGRDEA---NC
17 L33417      239  C-PTSEIQCGS--GECIHKKWRCDGDPDCKDGSDE---VNC
18 Q99087      148  C-NPAMFQCKDK-GICIPKLWACDGDPDCEDGSDE---EHC
                    *          *          *         *      **     *
```

b

```
 1 TMPRSS2-hum      41-134 VRLY-GPNFILQMYSSQRK-SWHPVCQDDWNENYGRAACRDMGYKNNFYSSQG---IVDDSG---STSFMKLNTSAGNV-----DIYKKLYH-SDACS--SKAVVSLRCLYH-SDACS--SKAVVSLRCL
 2 A48231-mouse     24-124 MGLVNGASANEGRVEIFYRGRWGTVCDNLWNLLDAHVVCRALGYENATQALGR--AAFGPGK---GPIMLDEVECTGTE-SSLASCRSLGWM-VSRCG--HEKDAGVVCSWM-VSRCG--HEKDAGVVCS
 3 D13381-rabbit   353-453 VRLVGGRGPHEGRVEILHNGQWGTVCDDHWELRAGQVVCRSLGYRGVKSVHKK--AYFGQGT---GPIWLNEVPCLGME-SSIEECKIRQWG-VRVCS--HGEDAGVTCTWG-VRVCS--HGEDAGVTCT
 4 P21757-hum      350-450 VRLVGGSGPHEGRVEILHSGQWGTICDDRWEVRVGQVVCRSLGYPGVQAVHKA--AHFGQGT---GPIWLNEVFCFGRE-SSIEECKIRQWG-TRACS--HSEDAGVTCTWG-TRACS--HSEDAGVTCT
 5 P21758-bovin    352-452 VRLVGGSGPHEGRVEIFHEGQWGTVCDDRWELRGGLVVCRSLGYKGVQSVHKR--AYFGKGT---GPIWLNEVFCFGKE-SSIEECRIRQWG-VRACS--HDEDAGVTCTWG-VRACS--HDEDAGVTCT
 6 P30204-mouse      9-109 VRLVGGSGAHEGRVEIFHQGQWGTICDDRWDIRAGQVVCRSLGYQEVLAVHKR--AHFGQGT---GPIWLNEVMCFGRE-SSIENCKINQWG-VLSCS--HSEDAGVTCTWG-VLSCS--HSEDAGVTCT
 7 P16264a-urchin   43-144 IRLIHGRTENEGSVEIYHATRWGGVCDWWWHMENANVTCKQLGFPGARQFYRR--AYFGAHV---TTFWVYKMNCLGNE-TRLEDCYHRPYGRPWLCN--AQWAAGVECLYGRPWLCN--AQWAAGVECL
 8 P16264b-urchin  153-256 LRMILGDVPNEGTLETFWDGAWGSVCHTDFGTPDGNVACRQMGYSRGVKSIKT-DGHFGFST---GPIILDAVDCEGTE-AHITECNMPVTPYQHACPYTHNWDVGVVCKTPYQHACPYTHNWDVGVVCK
 9 P16264c-urchin  264-366 IRLMDGSGPHEGRVEIWHDDAWGTICDDGWDWADANVVCRQAGYRGAVKASGFKGEDFGFTW---APIHTSFVMCTGVE-DRLIDCILRDGW-THSCY--HVEDASVVCAGW-THSCY--HVEDASVVCA
10 P16264d-urchin  382-485 VRIV-GMGQGQGRVEVSLGNGWGRVCDPDWSDHEAKTVCYHAGYKWGASRAAG-SAEVSAPFDLEAPFIIDGITCSGVENETLSQCQMKVSA-DMTCA---TGDVGVVCESA-DMTCA---TGDVGVVCE
                             *                 *   *   *       *  **                                    *                   *         *      *           *
```

FIG. 6. (**a**) Amino acid sequence comparison of the LDLRA domain of TMPRSS2 with a few selected such domains of other proteins: EK1, 4 (enterokinase—bovine); LDLR1–7 (LDL receptor class A domains—human); L34049a, b (LDL receptor-related protein 2 precursor, megalin—rat); U13637a, b (putative vitellogenin receptor precursor—*Drosophila melanogaster*); P07358 (complement C8 β chain precursor—human); U60975 (hybrid receptor gp250 precursor—human); L33417 (VLDL receptor precursor—mouse); and Q99087 (LDL receptor 1 precursor—*Xenopus*). (**b**) Amino acid sequence comparison of the SRCR domain of TMPRSS2 with a few selected such domains of other proteins: A48231 (cyclophilin C-associated protein precursor—mouse); D13381 (mRNA for macrophage scavenger receptor type I subunit—rabbit); P21757 (macrophage scavenger receptor type I and II—human); P21758 (macrophage scavenger receptor type I and II—bovine); P30204 (macrophage scavenger receptor type I and II—mouse); and P16264a–d (EGG peptide speract receptor precursor). (**c**) Amino acid sequence comparison of the protease domain of TMPRSS2 with a few other selected proteases: P00766 (chymotrypsinogen—bovine); P03952 (plasma kallikrein precursor—human); P05981 (serine protease hepsin—human); P07146 (trypsinogen precursor—mouse); P07477 (trypsinogen I precursor—human); P07478 (trypsinogen II precursor—human); P14272 (plasma kallikrein precursor—rat); P15157 (α-tryptase precursor—human); P17538 (chymotrypsinogen B precursor—human); P20231 (β-tryptase precursor—human); P20231 (β-tryptase precursor—human); P26262 (plasma kallikrein precursor—mouse); P98073 (enterokinase—human); Q05511 (serine protease hepsin—rat); X07002 (serine protease hepsin—human); X14844 (acrosin precursor—pig); and Y00970 (acrosin precursor—human).

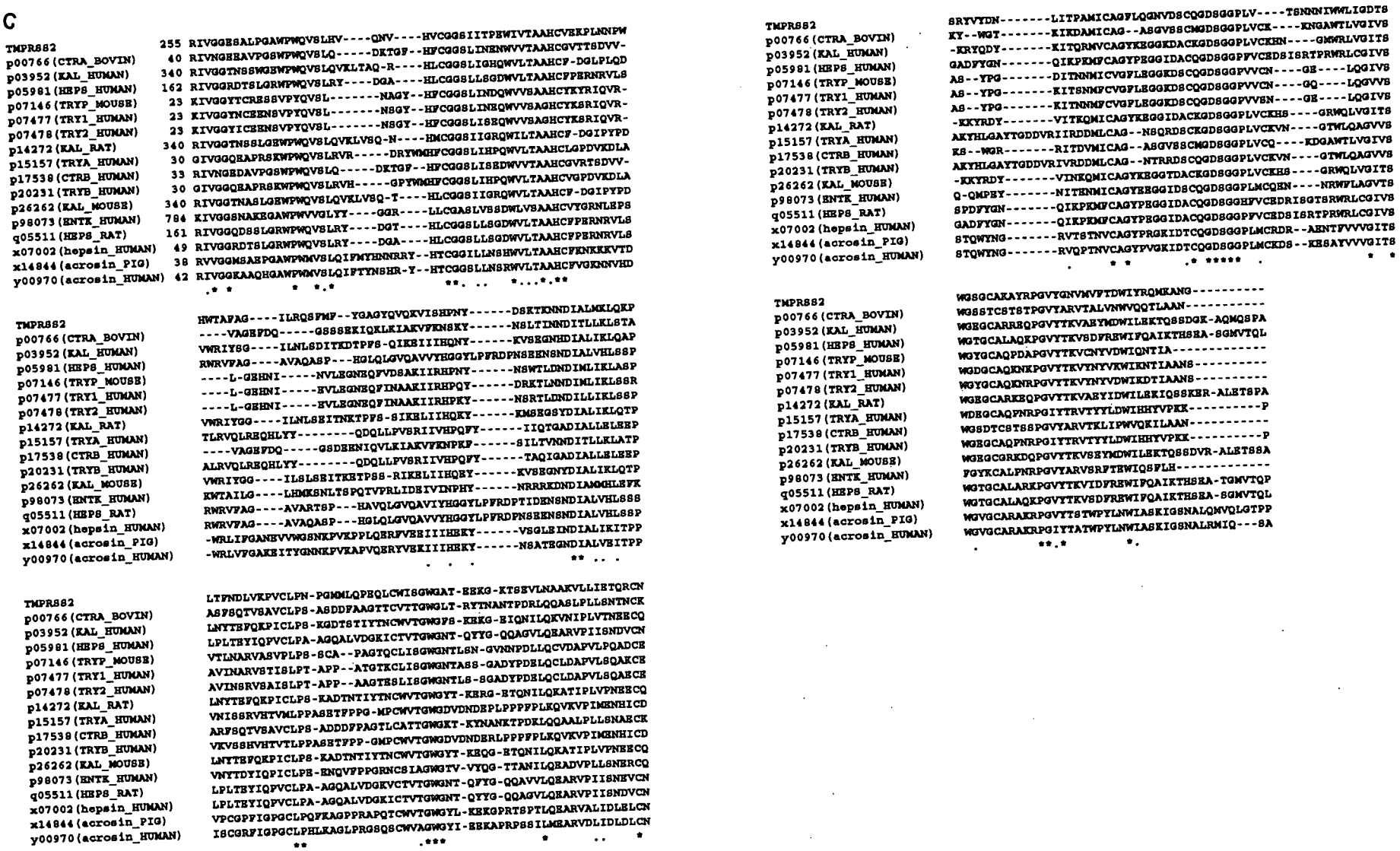
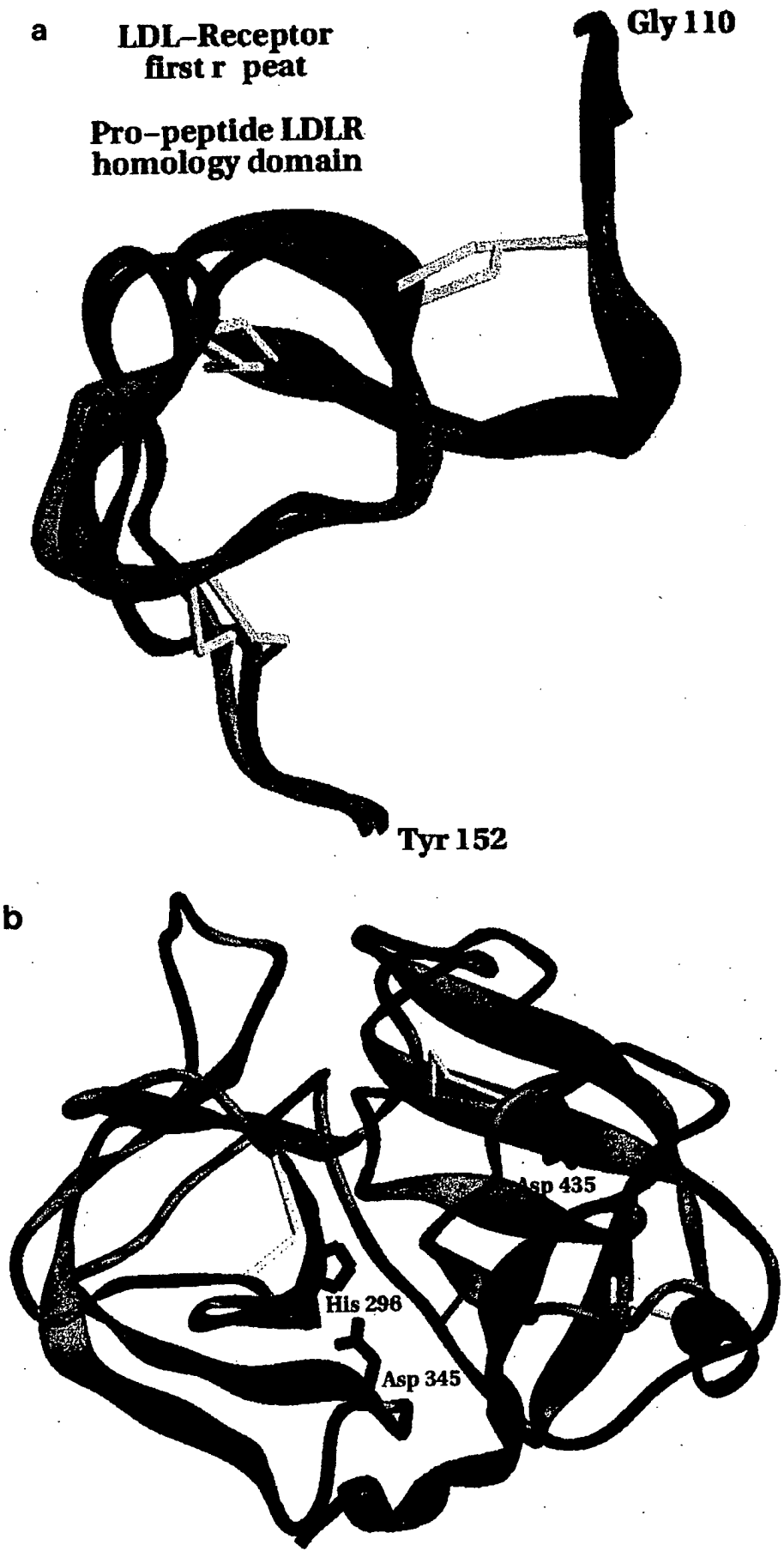


FIG. 6—*Continued*


a
LDL–Receptor
first r peat
Pro–peptide LDLR
homology domain
Gly 110
Tyr 152
b
Asp 435
His 296
Asp 345

1993), suggested that amino acids 84–106 of TMPRSS2 were hydrophobic and likely to be a transmembrane domain (Figs. 1 and 5). This hydrophobic sequence is not preceded by a recognizable leader sequence. These findings are compatible with a type II integral membrane protein in which the amino-terminus is at the cytoplasmic side of the membrane (Parks and Lamb, 1993). These features (a type II integral membrane polypeptide with an extracellular protease domain) are similar to those of mammalian hepsins (Leytus *et al.*, 1988; Tsuji *et al.*, 1991). This latter protein is important for cell growth and maintenance of normal cell morphology (Kurachi *et al.*, 1994); however, the underlying mechanisms for the biological activities are unknown.

LDLRA Domain

In addition to the transmembrane domain, TMPRSS2 contains a protein motif of the so-called LDLRA (low-density lipoprotein receptor A) domain extending from Cys113 to Cys148 (Figs. 1 and 5). This structural motif (PDOC00929; http://www.expasy.ch/cgi-bin/get-prodoc-entry?PDOC00929) was found in the low-density lipoprotein receptor gene, which contains seven successive such domains (Südhof *et al.*, 1985). A typical LDLRA domain is about 40 amino acids long and contains 6 disulfide-bound cysteines (cysteines 113, 120, 126, 133, 139, and 148 in TMPRSS2). Similar domains have been found in both extracellular and membrane proteins, including the VLDL receptor; gp330; *Drosophila* putative vitellogenin receptor; human enterokinase complement factor I; complement components C6, C7, C8, and C9; perlecan; PKD1; and vertebrate integral membrane protein DGCR2/IDD (Daly *et al.*, 1995). The amino acid comparison of the single LDLRA domain of TMPRSS2 with other similar domains is shown in Fig. 6a. The predicted 3D structure of this domain and its comparison with the first such domain of the LDLR is shown in Fig. 7a. The LDLRA domains form the binding site for LDL and calcium; the acidic residues between the fourth and the sixth cysteines are important for high affinity-binding of positively charged sequences in LDLR ligands (van Driel *et al.*, 1987; Mahley, 1988).

The SRCR Domain

An SRCR domain (Resnick *et al.*, 1994) was also identified in TMPRSS2 extending from Val149 to Leu242. SRCR domains are approximately 100 amino acids long and rich in cysteine. The overall consensus sequence derived from more than 40 such domains from different proteins revealed a consensus sequence at 41 of 101 residues (Resnick *et al.*, 1994). Two groups of SRCR domains are recognized, group A and group B, differing in the number of conserved cysteines. The SRCR domain of TMPRSS2 contains the pattern compatible with group A SRCR. The sequence homology to different examples of group A SRCR domains is shown in Fig. 6b. The SRCR domains were first found in type I macrophage scavenger receptor (Freeman *et al.*, 1990) but subsequently in many other sequences (for a comprehensive list, see Resnick *et al.*, 1994). The SRCR domain is reminiscent of but different from immunoglobulin domains. Proteins with SRCR domains are either at the cell surface or secreted into plasma or other body fluids. Some proteins such as the WC1 antigen or M130 contain nine or more such domains while others such as the MSR (macrophage scavenger receptor type I) and the secreted CF1 (complement factor 1) or cyclophilin C contain only one domain. The biochemical functions of the SRCR domain have not been established with certainty; however, most of these domains are involved with binding to the cell surface of extracellular molecules.

Protease Domain

The most striking feature of the TMPRSS2 predicted polypeptide is its similarity with members of serine protease family of proteins. The serine protease domain extends from amino acid residue Arg255 to the carboxyl-terminus of the predicted polypeptide. There is approximately 45–55% identity with several members of the serine protease family; the best similarities are with human hepsin (X07002), human enterokinase (P98073), and human kallikrein (P03952). The features of the protease domain of TMPRSS2 are compatible with the S1 family of the SA clan of serine-type peptidases as characterized by Rawlings and Barrett (1994). The prototype of this family is chymotrypsin and the 3D structure of some of its members has already been resolved. For a comprehensive list of the S1 serine-type peptidases see SWISS-PROT (http://www.expasy.ch/cgi-bin/lists?peptidas.txt). TMPRSS2 exhibits conservation of serine protease sequence motifs (Fig. 6c); in particular, the active site residues can be identified as His296, Asp345, and Ser441. TMPRSS2 is predicted to cleave after Lys or Arg residues since it contains Asp435 at the base of the specificity pocket (S1 subsite) that binds to the substrate. The predicted 3D structure of the protease domain of TMPRSS2 is shown in Fig. 7b. The protein model was built using the SWISS-MODEL server for automated comparative protein modeling (Peitsch, 1995, 1996) as described under Materials and Methods. It is of interest that TMPRSS2 is highly homologous to hepsin, another protease that contains a transmembrane domain and is thus a type II integral membrane protein with its protease domain

FIG. 7. (a) Ribbon model of the LDLRA domain of TMPRSS2. The NMR structure of the LDL receptor A domain is depicted in blue while the TMPRSS2 LDLRA homology domain is shown in red. The three disulfide bonds are shown in yellow. (b) Ribbon model of the protease domain of TMPRSS2. The full protein structure is depicted as a gray ribbon, while the active sites are shown with colored residues (His296, blue; Asp345, red; Ser441, green). The side chain of Asp435, which determines the Lys/Arg specificity of the TMPRSS2 protease, is shown in red. The three disulfide bonds are depicted in yellow, while two free cysteines are shown as orange bars.

in the extracellular space (Kurachi *et al.,* 1994; Leytus *et al.,* 1988; Tsuji *et al.,* 1991). TMPRSS2 contains nine conserved cysteine residues which by homology to other proteases most likely form the following intrasubunit disulfide bonds Cys826-Cys842, Cys926-Cys993, Cys957-Cys972, and Cys983-Cys1011 and the intersubunit disulfide bond involving Cys758-Cys912 which probably joins the catalytic protease subunit with the nonprotease part of the polypeptide. The protease domain does not contain potential N-glycosylation sites while the remainder of the predicted polypeptide contains two such potential sites (N213, in the SRCR domain, and N249). The amino-terminal Ile of the protease domain is preceded by Arg in the context of a peptide sequence Arg-Ile-Val-Gly-Gly (RIVGG), which is typical for the proteolytic activator site of many serine protease zymogens (Rawlings and Barrett, 1994). The potential cleavage between Arg and Ile, which would be similar to the activation mechanism of other serine protease zymogens, would convert TMPRSS2 to an activated form consisting of a nonprotease and a protease catalytic subunit linked by a disulfide bond that most probably involves Cys758 and Cys912.

DISCUSSION

In this paper we describe the cloning, chromosomal mapping, and initial characterization of a novel gene that maps on human chromosome 21q22.3 and encodes a polypeptide with multiple recognizable domains, namely LDLRA, SRCR, and serine protease domains. In addition, the presence of a transmembrane domain and the absence of a signal peptide suggest that this is a type II integral membrane protein. More biochemical experiments are necessary to further characterize the cellular localization of this protein and its physiological function. The biochemical events for the activation of the probable serine protease activity are unknown but are likely to be similar to those described above. It is of interest that the predicted TMPRSS2 protein contains additional domains (LDLRA and SRCR) that are potentially involved in binding with extracellular molecules or the cell surface. The molecules that are cleaved by or that bind to TMPRSS2 are unknown. There are several tissues that are shown by Northern blot analysis to express the TMPRSS2 gene. The site of the strongest expression is the small intestine; however, other tissues including heart, lung, and liver also showed a significant amount of TMPRSS2 mRNA. The function of this protein in these tissues remains elusive.

Are there any monogenic disorders associated with the TMPRSS2? Several monogenic phenotypes due to mutations in unknown genes have been mapped by linkage analysis to chromosome 21q22.3; these include APECED (Aaltonen *et al.,* 1994; OMIM 240300), an autoimmune disorder, two forms of autosomal recessive deafness (Bonné-Tamir *et al.,* 1996; Veske *et al.,* 1996; OMIM 601072); Knobloch syndrome (Sertie *et al.,* 1996; OMIM 267750); one locus for manic depressive illness (Smyth *et al.,* 1997; OMIM 125480); and one locus for holoprosencephaly (Muenke *et al.,* 1995; OMIM 236100). All of these phenotypes are mapped more distal to TMPRSS2, and it is therefore unlikely that TMPRSS2 is a candidate gene for any of these disorders.

Many human disorders are due to deficiency of other serine proteases. For example, deficiencies of coagulation factors such as Factor XII (OMIM 234000), Factor X (OMIM 227600), Factor IX (OMIM 306900), and Factor VII (OMIM 227500) belong to these disorders. Additional examples of such disorders are enterokinase deficiency (Hadorn *et al.,* 1969; OMIM 226200), trypsinogen deficiency (Townes, 1965; OMIM 276000), and hereditary pancreatitis due to mutations in the cationic trypsinogen gene (Whitcomb *et al.,* 1996). The generation of mice with targeted disruption of the mouse TMPRSS2 gene will enhance our understanding of the function of this gene and will provide candidate phenotypes for further investigation.

Is the overexpression of three copies of the TMPRSS2 involved in one of the phenotypes of Down syndrome? TMPRSS2 maps outside the so-called Down syndrome critical region (DSCR; between markers D21S17 and ETS2), triplication of which is associated with many phenotypes of Down syndrome (Delabar *et al.,* 1993). However, the existence of a single DSCR has recently been challenged since rare patients with proximal trisomy 21 not including the D21S17–ETS2 region displayed some of the phenotypes of Down syndrome (Korenberg *et al.,* 1994). In addition, a wider region from D21S17 to and including MX1 was associated with several phenotypes, including the heart defect and some dysmorphic features of the syndrome (Delabar *et al.,* 1993; Korenberg *et al.,* 1994). Since the TMPRSS2 gene is within this interval it is formally a candidate for some phenotype(s) of Down syndrome. Transgenic mice that overexpress the murine extracellular protein urokinase-type plasminogen activator have been shown to exhibit abnormal phenotypes (learning disabilities) (Meiri *et al.,* 1994). The study of transgenic mice that overexpress the murine homologue of the human TMPRSS2 gene may contribute to the understanding of the potential involvement of this gene in the pathogenesis of Down syndrome. A mouse model with partial trisomy 16 (which corresponds to a partial human trisomy 21 from APP to MX1) has recently been made (Reeves *et al.,* 1995). It would be of interest to know if the murine homologue of the TMPRSS2 gene is included in the triplicated part of mouse chromosome 16.

ACKNOWLEDGMENTS

We thank P. de Jong for the HC21-specific cosmid library LL2INCO2-Q, D. Patterson for the chromosome-21-specific somatic cell hybrids, and H. S. Scott for critically reading the manuscript. This study was supported by Grant 31-40500.94 from the Swiss FNRS, the European Union Grants GENE-CT93-0015 and PL-970302, and funds from the University and the Cantonal Hospital of Geneva.

REFERENCES

Aaltonen, J., Bjorses, P., Sandkuijl, L., Perheentupa, J., and Peltonen, L. (1994). An autosomal locus causing autoimmune disease: Autoimmune polyglandular disease type I assigned to chromosome 21. *Nature Genet.* **8:** 83–87.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Antonarakis, S. E. (1993). Human chromosome 21: Genome mapping and exploration, circa 1993. *Trends in Genet.* **9:** 142–148.

Bonné-Tamir, B., DeStefano, A. L., Briggs, C. E., Adair, R., Franklyn, B., Weiss, S., Korostishevsky, M., Frydman, M., Baldwin, C. T., and Farrer, L. A. (1996). Linkage of congenital recessive deafness (gene DFNB10) to chromosome 21q22.3. *Am. J. Hum. Genet.* **58:** 1254–1259.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991). Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* **88:** 4005–4009.

Chen, H., Chrast, R., Rossier, C., Morris, M. A., Lalioti, M. D., and Antonarakis, S. E. (1996). Cloning of 559 potential exons of genes of human chromosome 21 by exon trapping. *Genome Res.* **6:** 747–760.

Cheng, J. F., Boyartchuk, V., and Zhu, Y. W. (1994). Isolation and mapping of human chromosome 21 cDNA: Progress in constructing a chromosome 21 expression map. *Genomics* **23:** 75–84.

Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billaut, A., Guasconi, G., Gervy, P., LeGall, I., Soularue, P., Grinas, L., Bougueleret, L., Bellanne-Chantelot, C., Lacroix, B., Barillot, E., Gesnouin, P., Pook, S., Vaysseix, G., Frelat, G., Schmitz, A., Sambucy, J. L., Bosch, A., Estivill, X., Weissenbach, J., Vignal, A., Riethman, H., Cox, D., Patterson, D., Gardiner, K., Hattori, M., Sakaki, Y., Ichikawa, H., Ohki, M., Le Paslier, D., Heilig, R., Antonarakis, S. E., and Cohen, D. (1992). A continuum of overlapping clones spanning the entire chromosome 21q. *Nature* **359:** 380–386.

Church, D. M., Stotler, C. J., Rutter, J. L., Murrell, J. R., Trofatter, J. A., and Buckler, A. J. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6:** 98–105.

Daly, N. L., Scanlon, M. J., Djordjevic, J. T., Kroon, P. A., and Smith, R. (1995). Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. *Proc. Natl. Acad. Sci. USA* **92:** 63334–6338.

Delabar, J. M., Theophile, D., Rahmani, Z., Chettouh, Z., Blouin, J. L., Prieur, M., Noël, B., and Sinet, P. M. (1993). Molecular mapping of twenty-four features of Down syndrome on chromosome 21. *Eur. J. Hum. Genet.* **1:** 114–124.

Drwinga, H. L., Toji, L. H., Kim, C. H., Greene, A. E., and Mulivor, R. A. (1993). NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. *Genomics* **16:** 311–314.

Epstein, C. J. (1989). Down syndrome, trisomy 21. *In* "The Metabolic Basis of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, Eds.), pp. 291–326, McGraw–Hill, New York.

Freeman, M., Ashkenas, J., Rees, D. J. G., Kingsley, D. M., Copeland, N. G., Jenkins, N. A., and Krieger, M. (1990). An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II-murine macrophage scavenger receptors. *Proc. Natl. Acad. Sci. USA* **87:** 8810–8814.

Hadorn, B., Tarlow, M. J., Lloyd, J. K., and Wolff, O. H. (1969). Intestinal enterokinase deficiency. *Lancet* **I:** 812–813.

Hofmann, K., and Stoffel, W. (1993). Tmbase—A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **347:** 166.

Korenberg, J. R., Chen, X. N., Schipper, R., Sun, Z., Gonsky, R., Gerwehr, S., Carpenter, N., Daumer, D., Dignan, P., Disteche, C., Graham, J. M., Hugdins, L., McGillivray, B., Miyazaki, K., Ogasawara, N., Park, J. P., Pagon, R., Pueschel, S., Sack, G., Say, B., Schuffenhauer, S., Soukup, S., and Yamanaka, T. (1994). Down syndrome phenotype: The consequences of chromosomal imbalance. *Proc. Natl. Acad. Sci. USA* **91:** 4997–5001.

Kurachi, K., Torres-Rosado, A., and Tsuji, A. (1994). Hepsin. *Methods Enzymol.* **244:** 100–114.

Leytus, S. P., Loeb, K. R., Hagen, S. F., Kurachi, K., and Davie, E. W. (1988). A novel trypsin-like serine protease (Hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells. *Biochemistry* **27:** 1067–1074.

Lucente, D., Chen, H. M., Shea, D., Samec, S. N., Rutter, M., Chrast, R., Rossier, C., Buckler, A., Antonarakis, S. E., and McCormick, M. K. (1995). Localization of 102 exons to a 2.5 Mb region of chromosome 21 involved in Down syndrome. *Hum. Mol. Genet.* **4:** 1305–1311.

Lüthy, R., Bowie, J. U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* **356:** 83–85.

Mahley, R. W. (1988). Apolipoprotein E: Cholesterol transport protein with expanding role in cell biology. *Science* **240:** 622–630.

Martin, C. H., Bondoc, M. M., Chiang, A., Cloutier, T., Davis, C. A., Ericsson, C. L., Jaklevic, M. A., Kim, R. J., Lee, M. T., Li, M., Mayeda, C. A., Steiert-El Kheir, A., and Palazzolo, M. J. (1994). Sequencing of the MX1 region of human chromosome 21. [Unpublished] [http://www2.ncbi.nlm.nih.gov/cgi-bin/genbank?L35675]

Meiri, N., Masos, T., Rosenblum, K., Miskin, R., and Dudai, Y. (1994). Overexpression of urokinase-type plasminogen activator in transgenic mice is correlated with impaired learning. *Proc. Natl. Acad. Sci. USA* **91:** 3196–3200.

Muenke, M., Bone, L. J., Mitchell, H. F., Hart, I., Walton, K., Hall-Johnson, K., Ippel, E. F., Dietz-Band, J., Kvaloy, K., Fan, C.-M., Tessier-Lavigne, M., and Patterson, D. (1995). Physical mapping of the holoprosencephaly critical region in 21q22.3, exclusion of SIM2 as a candidate gene for holoprosencephaly, and mapping of SIM2 to a region of chromosome 21 important for Down syndrome. *Am. J. Hum. Genet.* **57:** 10747–1079.

Parks, G. D., and Lamb, R. A. (1993). Role of NH_2-terminal positively charged residues in establishing membrane protein topology. *J. Biol. Chem.* **268:** 19101–19109.

Patterson, D., Rahmani, Z., Donaldson, D., Gardiner, K., and Jones, C. (1993). Physical mapping of chromosome 21. *Prog. Clin. Biol. Res.* **384:** 33–50.

Peitsch, M. C. (1995). Protein modelling by e-mail. *Bio/Technology* **13:** 658–660.

Peitsch, M. C. (1996). ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.* **24:** 274–279.

Peterson, A., Patil, N., Robbins, C., Wang, L., Cox, D. R., and Myers, R. M. (1994). A transcript map of the Down syndrome critical region on chromosome 21. *Hum. Mol. Genet.* **3:** 1735–1742.

Rawlings, N. D., and Barrett, A. J. (1994). Families of cysteine peptidases. *Methods Enzymol.* **244:** 19–61.

Reeves, R. H., Irving, N. G., Moran, T. H., Wohn, A., Kitt, C., Sisodia, S. S., Schmidt, C., Bronson, R. T., and Davisson, M. T. (1995). A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nature Genet.* **11:** 177–184.

Resnick, D., Pearson, A., and Krieger, M. (1994). The SRCR superfamily: A family reminiscent of the Ig superfamily. *Trends Biochem. Sci.* **19:** 5–8.

Sertie, A. L., Quimby, M., Moreira, E. S., Murray, J., Zatz, M., Antonarakis, S. E., and Passos-Bueno, M. R. (1996). A gene which causes severe ocular alterations and occipital encephalocele (Knobloch syndrome) is mapped to 21q22.3. *Hum. Mol. Genet.* **5:** 843–847.

Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins Struct. Funct. Genet.* **17:** 355–362.

Smyth, C., Kalsi, G., Curtis, D., Brynjolfsson J., O'Neill, J., Rifkin, L., Moloney, E., Murphy, P., Petursson, H., and Gurling, H. (1997). Two-locus admixture linkage analysis of bipolar and unipolar af-

fective disorder supports the presence of susceptibility loci on chromosomes 11p15 and 21q22. *Genomics* **39:** 271–278.

Südhof, T. C., Goldstein, J. L., Brown, M. S., and Russel, D. W. (1985). The LDL receptor gene: A mosaic of exons shared with different proteins. *Science* **228:** 815–822.

Tassone, F., Xu, N. X., Wade, H., Weissman, S., and Gardiner, K. (1994). High density transcriptional mapping of chromosome 21 by hybridization selection. *Am. J. Hum. Genet.* **55:** 272A.

Townes, P. L. (1965). Trypsinogen deficiency disease. *J. Pediat.* **66:** 275–285.

Tsuji, A., Torres-Rosado, A., Arai, T., Le Beau, M. M., Lemons, R. S., Chou, S. H., and Kurachi, K. (1991). *J. Biol. Chem.* **266:** 16948–16953.

van Driel, I. R., Goldstein, J. L., Südhof, T. C., and Brown, M. S. (1987). First cysteine-rich repeat in ligand-binding domain of low density lipoprotein receptor binds Ca^{2+} and monoclonal antibodies, but not lipoproteins. *J. Biol. Chem.* **262:** 17443–17449.

Veske, A., Oehlmann, R., Younus, F., Mohyuddin, A., Muller-Myhsok, B., Mehdi, S. Q., and Gal, A. (1996). Autosomal recessive nonsyndromic deafness locus (DFNB8) maps on chromosome 21q22 in a large consanguineous kindred from Pakistan. *Hum. Mol. Genet.* **5:** 165–168.

Whitcomb, D. C., Gorry, M. C., Preston, R. A., Furey, W., Sossenheimer, M. J., Ulrich, C. D., Martin, S. P., Gates, L. K., Jr., Amann, S. T., Toskes, P. P., Liddle, R., McGrath, K., Uomo, G., Post, J. C., and Ehrlich, G. D. (1996). Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nature Genet.* **14:** 141–145.